SEQUENCE LISTING

(1)INFORMATION FOR SEQ ID NO:1:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:756
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:1:

```
Pro Ala Ser Thr Tyr Arg Leu Gln Ile Ser Ala Glu Phe Thr Leu Phe
 1               5                  10                  15
Asp Ala Ala Arg Ile Val Pro Tyr Leu His Arg Leu Gly Ala Asp Trp
            20                  25                  30
Leu Tyr Leu Ser Pro Leu Leu Glu Ser Glu Ser Gly Ser Ser His Gly
        35                  40                  45
Tyr Asp Val Val Asp His Ser Arg Val Asp Ala Ala Arg Gly Gly Pro
    50                  55                  60
Glu Gly Leu Ala Glu Leu Ser Arg Ala Ala His Glu Arg Gly Met Gly
65                  70                  75                  80
Val Val Val Asp Ile Val Pro Asn His Val Gly Val Ala Thr Pro Lys
                85                  90                  95
Ala Asn Arg Trp Trp Trp Asp Val Leu Ala Arg Gly Gln Arg Ser Glu
            100                 105                 110
Tyr Ala Asp Tyr Phe Asp Ile Asp Trp Glu Phe Gly Gly Gly Arg Leu
        115                 120                 125
Arg Leu Pro Val Leu Gly Asp Gly Pro Asp Glu Leu Asp Ala Leu Arg
    130                 135                 140
Val Asp Gly Asp Glu Leu Val Tyr Tyr Glu His Arg Phe Pro Ile Ala
145                 150                 155                 160
Glu Gly Thr Gly Gly Gly Thr Pro Arg Glu Val His Asp Arg Gln His
                165                 170                 175
Tyr Glu Leu Met Ser Trp Arg Arg Ala Asp His Asp Leu Asn Tyr Arg
            180                 185                 190
Arg Phe Phe Ala Val Asn Thr Leu Ala Ala Val Arg Val Glu Asp Pro
        195                 200                 205
Arg Val Phe Asp Asp Thr His Arg Glu Ile Gly Arg Trp Ile Ala Glu
    210                 215                 220
Gly Leu Val Asp Gly Leu Arg Val Asp His Pro Asp Gly Leu Arg Ala
225                 230                 235                 240
Pro Gly Asp Tyr Leu Arg Arg Leu Ala Glu Leu Ala Gln Gly Arg Pro
                245                 250                 255
Ile Trp Val Glu Lys Ile Ile Glu Gly Asp Glu Arg Met Pro Pro Gln
            260                 265                 270
Trp Pro Ile Ala Gly Thr Thr Gly Tyr Asp Ala Leu Ala Gly Ile Asp
        275                 280                 285
Arg Val Leu Val Asp Pro Ala Gly Glu His Pro Leu Thr Gln Ile Val
    290                 295                 300
Asp Glu Ala Ala Gly Ser Pro Arg Arg Trp Ala Glu Leu Val Pro Glu
305                 310                 315                 320
Arg Lys Arg Ala Val Ala Arg Gly Ile Leu Asn Ser Glu Ile Arg Arg
                325                 330                 335
Val Ala Arg Glu Leu Gly Glu Val Ala Gly Asp Val Glu Asp Ala Leu
            340                 345                 350
Val Glu Ile Ala Ala Ala Leu Ser Val Tyr Arg Ser Tyr Leu Pro Phe
```

```
        355                 360                 365
Gly Arg Glu His Leu Asp Glu Ala Val Ala Ala Ala Gln Ala Ala Ala
    370                 375                 380
Pro Gln Leu Glu Ala Asp Leu Ala Ala Val Gly Ala Ala Leu Ala Asp
385                 390                 395                 400
Pro Gly Asn Pro Ala Ala Leu Arg Phe Gln Gln Thr Ser Gly Met Ile
                405                 410                 415
Met Ala Lys Gly Val Glu Asp Asn Ala Phe Tyr Arg Tyr Pro Arg Leu
            420                 425                 430
Thr Ser Leu Thr Glu Val Gly Gly Asp Pro Ser Leu Phe Ala Ile Asp
        435                 440                 445
Ala Ala Ala Phe His Ala Ala Gln Arg Asp Arg Ala Ala Arg Leu Pro
    450                 455                 460
Glu Ser Met Thr Thr Leu Thr Thr His Asp Thr Lys Arg Ser Glu Asp
465                 470                 475                 480
Thr Arg Ala Arg Ile Thr Ala Leu Ala Glu Ala Pro Glu Arg Trp Arg
                485                 490                 495
Arg Phe Leu Thr Glu Val Gly Gly Leu Ile Gly Thr Gly Asp Arg Val
            500                 505                 510
Leu Glu Asn Leu Ile Trp Gln Ala Ile Val Gly Ala Trp Pro Ala Ser
        515                 520                 525
Arg Glu Arg Leu Glu Ala Tyr Ala Leu Lys Ala Ala Arg Glu Ala Gly
    530                 535                 540
Glu Ser Thr Asp Trp Ile Asp Gly Asp Pro Ala Phe Glu Glu Arg Leu
545                 550                 555                 560
Thr Arg Leu Val Thr Val Ala Val Glu Glu Pro Leu Val His Glu Leu
                565                 570                 575
Leu Glu Arg Leu Val Asp Glu Leu Thr Ala Ala Gly Tyr Ser Asn Gly
            580                 585                 590
Leu Ala Ala Lys Leu Leu Gln Leu Leu Ala Pro Gly Thr Pro Asp Val
        595                 600                 605
Tyr Gln Gly Thr Glu Arg Trp Asp Arg Ser Leu Val Asp Pro Asp Asn
    610                 615                 620
Arg Arg Pro Val Asp Phe Ala Ala Ala Ser Glu Leu Leu Asp Arg Leu
625                 630                 635                 640
Asp Gly Gly Trp Arg Pro Pro Val Asp Glu Thr Gly Ala Val Lys Thr
                645                 650                 655
Leu Val Val Ser Arg Ala Leu Arg Leu Arg Arg Asp Arg Pro Glu Leu
            660                 665                 670
Phe Thr Ala Tyr His Pro Val Thr Ala Arg Gly Ala Gln Ala Glu His
        675                 680                 685
Leu Ile Gly Phe Asp Arg Gly Gly Ala Ile Ala Leu Ala Thr Arg Leu
    690                 695                 700
Pro Leu Gly Leu Ala Ala Ala Gly Gly Trp Gly Asp Thr Val Val Asp
705                 710                 715                 720
Val Gly Glu Arg Ser Leu Arg Asp Glu Leu Thr Gly Arg Glu Ala Arg
                725                 730                 735
Gly Ala Ala Arg Val Ala Glu Leu Phe Ala Asp Tyr Pro Val Ala Leu
            740                 745                 750
Leu Val Glu Thr
        755
```

(2)INFORMATION FOR SEQ ID NO:2:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:6

(B)TYPE:amino acid

(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:2:

Asp Ile Val Pro Asn His
1 5

(3)INFORMATION FOR SEQ ID NO:3:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:6
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:3:

Gly Thr Thr Gly Tyr Asp
1 5

(4)INFORMATION FOR SEQ ID NO:4:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:N-terminal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:4:

Pro Ala Ser Thr Tyr Arg Leu Gln Ile Ser Ala Glu Phe Thr Leu Phe
1 5 10 15
Asp Ala Ala Arg
20

(5)INFORMATION FOR SEQ ID NO:5:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:5:

Ser Leu Val Asp Pro Asp Asn Arg Arg Pro Val Asp Phe Ala Ala Ala
1 5 10 15
Ser Glu Leu Leu
20

(6)INFORMATION FOR SEQ ID NO:6:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:6:

Ala Asn Arg Trp Trp Trp Asp Val Leu Ala Arg Gly Gln Arg Ser Glu
1 5 10 15
Tyr Ala Asp Tyr
20

(7)INFORMATION FOR SEQ ID NO:7:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:2268
(B)TYPE:nucleic acid
(C)strandedness:double
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:7:

CCCGCCAGTA CCTACCGCCT TCAGATCTCG GCGGAGTTCA CCCTCTTCGA CGCGGCGCGC 60
ATCGTGCCCT ACCTGCACCG CCTCGGCGCC GACTGGCTGT ACCTCTCGCC GCTGCTCGAG 120
TCCGAGTCGG GCTCCTCGCA CGGCTACGAC GTGGTCGACC ACTCCCGCGT CGACGCCGCC 180
CGCGGCGGGC CGGAGGGGCT CGCCGAGCTC TCCCGTGCGG CGCACGAGCG CGGCATGGGC 240
GTCGTCGTCG ACATCGTGCC CAACCACGTC GGCGTCGCGA CGCCGAAGGC GAACCGCTGG 300
TGGTGGGACG TTCTGGCCCG TGGACAGCGG TCGGAGTACG CCGACTACTT CGACATCGAC 360
TGGGAGTTCG GCGGCGGCAG GCTGCGCCTG CCCGTGCTCG GCGACGGCCC CGACGAGCTC 420
GACGCGCTGA GAGTGGATGG CGACGAGCTC GTCTACTACG AGCACCGCTT CCCGATCGCC 480
GAGGGCACCG GCGGCGGCAC CCCGCGCGAG GTGCACGACC GGCAGCACTA CGAGCTGATG 540
TCGTGGCGGC GGGCCGACCA CGACCTCAAC TACCGCCGCT TCTTCGCCGT GAACACGCTC 600
GCCGCCGTAC GCGTCGAAGA CCCGCGCGTG TTCGACGACA CCCACCGCGA GATCGGCCGC 660
TGGATCGCCG AGGGCCTCGT CGACGGCCTG CGCGTCGACC ACCCCGACGG GCTGCGCGCC 720
CCCGGCGACT ACCTGCGCCG TCTCGCCGAG CTCGCCCAAG GCAGGCCGAT CTGGGTCGAG 780
AAGATCATCG AGGGCGACGA GCGGATGCCC CCGCAGTGGC CCATCGCCGG CACCACCGGC 840
TACGACGCGC TGGCCGGGAT CGACCGGGTG CTCGTCGACC CCGCGGGCGA GCATCCGCTC 900
ACCCAGATCG TCGACGAGGC GGCAGGCAGC CCCCGGCGCT GGGCCGAGCT GGTTCCCGAG 960
CGCAAGCGGG CCGTCGCCCG CGGCATCCTG AACTCCGAGA TCCGCCGCGT CGCCCGCGAA 1020
CTCGGAGAGG TCGCCGGCGA CGTCGAAGAC GCGCTCGTCG AGATCGCCGC CGCCCTGTCC 1080
GTCTACCGCA GCTACCTGCC GTTCGGGCGC GAGCACCTCG ACGAAGCCGT GGCCGCCGCG 1140
CAGGCCGCAG CCCCCCAGCT CGAGGCCGAC CTCGCCGCCG TCGGCGCAGC GCTCGCCGAC 1200
CCGGGCAACC CCGCCGCGCT CCGCTTCCAG CAGACCAGCG GCATGATCAT GGCCAAGGGC 1260
GTCGAGGACA ACGCGTTCTA CCGCTACCCC CGGCTCACCT CGCTGACCGA GGTCGGGGGA 1320
GACCCGAGCC TGTTCGCGAT CGACGCGGCC GCCTTCCACG CGGCGCAGCG CGACCGCGCC 1380
GCCCGGCTGC CCGAGTCGAT GACGACGCTG ACCACCCACG ACACCAAGCG CAGCGAAGAC 1440
ACCCGGGCGC GGATCACCGC GCTCGCCGAG GCCCCCGAAC GCTGGCGGCG CTTCCTGACC 1500
GAGGTCGGCG GGCTCATCGG AACGGGCGAC CGGGTGCTGG AGAACCTGAT CTGGCAGGCG 1560
ATCGTCGGCG CGTGGCCGGC GAGCCGGGAG CGGCTCGAGG CCTACGCGCT GAAGGCCGCG 1620
CGCGAAGCCG GCGAGTCGAC CGACTGGATC GACGGCGACC CCGCGTTCGA AGAGCGGCTG 1680
ACCCGCCTGG TCACGGTCGC CGTCGAGGAG CCGCTCGTGC ACGAGCTGCT CGAGCGGCTC 1740
GTCGACGAGC TGACGGCGGC CGGGTACTCC AACGGCCTCG CGGCGAAGCT GCTGCAGCTG 1800
CTCGCCCCCG GAACCCCCGA CGTGTACCAG GGCACGGAAC GCTGGGACCG GTCGCTGGTG 1860
GACCCGGACA ACCGTCGCCC CGTGGATTTC GCCGCGGCAT CCGAGCTGCT CGACCGCCTC 1920
GACGGCGGCT GGCGGCCGCC CGTCGACGAG ACCGGCGCGG TCAAGACGCT CGTCGTCTCC 1980
CGCGCGCTGC GGCTGCGCCG CGACCGGCCC GAGCTGTTCA CCGCGTACCA CCCGGTCACG 2040
GCGCGCGGCG CGCAGGCCGA GCACCTGATC GGCTTCGACC GCGGCGGCGC GATCGCCCTG 2100
GCCACCCGCC TGCCGCTCGG CCTCGCCGCC GCAGGCGGCT GGGGCGACAC GGTCGTCGAC 2160
GTCGGCGAGC GGAGCCTGCG CGACGAGCTG ACCGGCCGCG AGGCCCGCGG AGCGGCGCGC 2220
GTGGCCGAGT TGTTCGCCGA CTACCCCGTC GCCCTGCTGG TGGAGACA 2268

(8)INFORMATION FOR SEQ ID NO:8:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:28
(B)TYPE:nucleic acid

(C)strandedness:double
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:8:

TTTTTTAATA AAATCAGGAG GAAAAAAT

(9)INFORMATION FOR SEQ ID NO:9:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:575
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:9:

```
Met Asn Arg Arg Phe Pro Val Trp Ala Pro Gln Ala Ala Gln Val Thr
 1               5                  10                  15
Leu Val Val Gly Gln Gly Arg Ala Glu Leu Pro Leu Thr Arg Asp Glu
            20                  25                  30
Asn Gly Trp Trp Ala Leu Gln Gln Pro Trp Asp Gly Gly Pro Asp Leu
        35                  40                  45
Val Asp Tyr Gly Tyr Leu Val Asp Gly Lys Gly Pro Phe Ala Asp Pro
    50                  55                  60
Arg Ser Leu Arg Gln Pro Arg Gly Val His Glu Leu Gly Arg Glu Phe
65                  70                  75                  80
Asp Pro Ala Arg Tyr Ala Trp Gly Asp Asp Gly Trp Arg Gly Arg Asp
                85                  90                  95
Leu Thr Gly Ala Val Ile Tyr Glu Leu His Val Gly Thr Phe Thr Pro
            100                 105                 110
Glu Gly Thr Leu Asp Ser Ala Ile Arg Arg Leu Asp His Leu Val Arg
        115                 120                 125
Leu Gly Val Asp Ala Val Glu Leu Leu Pro Val Asn Ala Phe Asn Gly
    130                 135                 140
Thr His Gly Trp Gly Tyr Asp Gly Val Leu Trp Tyr Ala Val His Glu
145                 150                 155                 160
Pro Tyr Gly Gly Pro Glu Ala Tyr Gln Arg Phe Val Asp Ala Cys His
                165                 170                 175
Ala Arg Gly Leu Ala Val Val Gln Asp Val Val Tyr Asn His Leu Gly
            180                 185                 190
Pro Ser Gly Asn His Leu Pro Asp Phe Gly Pro Tyr Leu Gly Ser Gly
        195                 200                 205
Ala Ala Asn Thr Trp Gly Asp Ala Leu Asn Leu Asp Gly Pro Leu Ser
    210                 215                 220
Asp Glu Val Arg Arg Tyr Ile Ile Asp Asn Ala Val Tyr Trp Leu Arg
225                 230                 235                 240
Asp Met His Ala Asp Gly Leu Arg Leu Asp Ala Val His Ala Leu Arg
                245                 250                 255
Asp Ala Arg Ala Leu His Leu Leu Glu Glu Leu Ala Ala Arg Val Asp
            260                 265                 270
Glu Leu Ala Gly Glu Leu Gly Arg Pro Leu Thr Leu Ile Ala Glu Ser
        275                 280                 285
Asp Leu Asn Asp Pro Lys Leu Ile Arg Ser Arg Ala Ala His Gly Tyr
    290                 295                 300
Gly Leu Asp Ala Gln Trp Asp Asp Asp Val His His Ala Val His Ala
305                 310                 315                 320
Asn Val Thr Gly Glu Thr Val Gly Tyr Tyr Ala Asp Phe Gly Gly Leu
                325                 330                 335
```

```
Gly Ala Leu Val Lys Val Phe Gln Arg Gly Trp Phe His Asp Gly Thr
            340                 345                 350
Trp Ser Ser Phe Arg Glu Arg His His Gly Arg Pro Leu Asp Pro Asp
        355                 360                 365
Ile Pro Phe Arg Arg Leu Val Ala Phe Ala Gln Asp His Asp Gln Val
    370                 375                 380
Gly Asn Arg Ala Val Gly Asp Arg Met Ser Ala Gln Val Gly Glu Gly
385                 390                 395                 400
Ser Leu Ala Ala Ala Ala Ala Leu Val Leu Leu Gly Pro Phe Thr Pro
                405                 410                 415
Met Leu Phe Met Gly Glu Glu Trp Gly Ala Arg Thr Pro Trp Gln Phe
            420                 425                 430
Phe Thr Ser His Pro Glu Pro Glu Leu Gly Glu Ala Thr Ala Arg Gly
        435                 440                 445
Arg Ile Ala Glu Phe Ala Arg Met Gly Trp Asp Pro Ala Val Val Pro
    450                 455                 460
Asp Pro Gln Asp Pro Ala Thr Phe Ala Arg Ser His Leu Asp Trp Ser
465                 470                 475                 480
Glu Pro Glu Arg Glu Pro His Ala Gly Leu Leu Ala Phe Tyr Thr Asp
                485                 490                 495
Leu Ile Ala Leu Arg Arg Glu Leu Pro Val Asp Ala Pro Ala Arg Glu
            500                 505                 510
Val Asp Ala Asp Glu Ala Arg Gly Val Phe Ala Phe Ser Arg Gly Pro
        515                 520                 525
Leu Arg Val Thr Val Ala Leu Arg Pro Gly Pro Val Gly Val Pro Glu
    530                 535                 540
His Gly Gly Leu Val Leu Ala Tyr Gly Glu Val Arg Ala Gly Ala Ala
545                 550                 555                 560
Gly Leu His Leu Asp Gly Pro Gly Ala Ala Ile Val Arg Leu Glu
                565                 570                 575
```

(10)INFORMATION FOR SEQ ID NO:10:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:6

(B)TYPE:amino acid

(D)TOPOLOGY:linear

(ii)MOLECULE TYPE:peptide

(v)FRAGMENT TYPE:internal fragment

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:10:

```
Trp Gly Tyr Asp Gly Val
  1               5
```

(11)INFORMATION FOR SEQ ID NO:11:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:6

(B)TYPE:amino acid

(D)TOPOLOGY:linear

(ii)MOLECULE TYPE:peptide

(v)FRAGMENT TYPE:internal fragment

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:11:

```
Asp Val Val Tyr Asn His
  1               5
```

(12)INFORMATION FOR SEQ ID NO:12:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:7
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:12:

Arg Leu Asp Ala Val His Ala
1 5

(13)INFORMATION FOR SEQ ID NO:13:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:7
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:13:

Ile Ala Glu Ser Asp Leu Asn
1 5

(14)INFORMATION FOR SEQ ID NO:14:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:N-terminal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:14:

Met Asn Arg Arg Phe Pro Val Trp Ala Pro Gln Ala Ala Gln Val Thr
1 5 10 15
Leu Val Val Gly
20

(15)INFORMATION FOR SEQ ID NO:15:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear
(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:15:

Ser Arg Ala Ala His Gly Tyr Gly Leu Asp Ala Gln Trp Asp Asp Asp
1 5 10 15
Val His His Ala
20

(16)INFORMATION FOR SEQ ID NO:16:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:20
(B)TYPE:amino acid
(D)TOPOLOGY:linear

(ii)MOLECULE TYPE:peptide
(v)FRAGMENT TYPE:internal fragment
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:16:

```
Asp Glu Asn Gly Trp Trp Ala Leu Gln Gln Pro Trp Asp Gly Gly Pro
 1               5                  10                  15
Asp Leu Val Asp
            20
```

(17)INFORMATION FOR SEQ ID NO:17:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:1725
(B)TYPE:nucleic acid
(C)strandedness:double
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:17:

```
ATGAACCGAC GATTCCCGGT CTGGGCGCCC CAGGCCGCGC AGGTGACGCT CGTCGTGGGC   60
CAAGGCCGCG CCGAACTCCC GCTGACCCGC GACGAGAACG GATGGTGGGC TCTTCAGCAG  120
CCGTGGGACG GCGGCCCCGA CCTCGTCGAC TACGGCTACC TCGTCGACGG CAAGGGCCCC  180
TTCGCCGACC CGCGGTCGCT GCGGCAGCCG CGCGGCGTGC ACGAGCTCGG CCGCGAATTC  240
GACCCCGCCC GCTACGCGTG GGGCGACGAC GGATGGCGCG GCCGAGACCT CACCGGAGCC  300
GTGATCTACG AACTGCACGT CGGCACCTTC ACCCCTGAGG GAACGCTGGA CAGCGCCATC  360
CGTCGCCTCG ACCACCTGGT GCGCCTCGGC GTCGACGCGG TCGAGCTGCT GCCCGTCAAC  420
GCGTTCAACG GCACCCACGG CTGGGGCTAC GACGGGGTGC TCTGGTACGC GGTGCACGAG  480
CCCTACGGCG GCCCGGAGGC GTACCAGCGC TTCGTCGACG CCTGCCACGC CCGCGGCCTC  540
GCCGTCGTGC AGGACGTCGT CTACAACCAC CTGGGCCCGA GCGGCAACCA CCTGCCCGAC  600
TTCGGCCCCT ACCTCGGGTC GGGCGCCGCC AACACCTGGG GCGACGCGCT GAACCTCGAC  660
GGGCCGCTCT CCGACGAGGT GCGGCGGTAC ATCATCGACA ACGCGGTGTA CTGGCTGCGC  720
GACATGCACG CCGACGGGCT GCGGCTCGAC GCCGTGCACG CGCTGCGCGA CGCCCGCGCG  780
CTGCACCTGC TCGAAGAGCT CGCCGCCCGC GTCGACGAGC TGGCGGGCGA GCTCGGCCGG  840
CCGCTGACGC TCATCGCCGA GAGCGACCTG AACGACCCGA AGCTGATCCG CTCCCGCGCG  900
GCGCACGGCT ACGGCCTCGA CGCCCAGTGG GACGACGACG TGCACCACGC GGTGCACGCC  960
AACGTGACCG GCGAGACCGT CGGCTACTAC GCCGACTTCG GCGGGCTCGG CGCCCTCGTC 1020
AAGGTGTTCC AGCGCGGCTG GTTCCACGAC GGCACCTGGT CGAGCTTCCG CGAGCGGCAC 1080
CACGGCCGGC CGCTCGACCC CGACATCCCG TTCCGCCGGC TCGTCGCCTT CGCGCAGGAT 1140
CACGACCAGG TCGGCAACCG AGCGGTCGGC GACCGCATGT CGGCGCAGGT CGGCGAGGGT 1200
TCGCTCGCCG CCGCGGCGGC GCTCGTGCTG CTCGGCCCGT TCACCCCGAT GCTGTTCATG 1260
GGCGAGGAGT GGGGCGCGCG CACCCCGTGG CAGTTCTTCA CCTCCCACCC CGAGCCCGAG 1320
CTGGGGGAGG CGACGGCGCG CGGGCGCATC GCCGAGTTCG CCCGCATGGG CTGGGACCCG 1380
GCAGTCGTGC CCGACCCGCA GGACCCGGCC ACCTTCGCCC GCTCGCACCT GGACTGGTCC 1440
GAGCCCGAGC GGGAACCGCA CGCGGGCCTG CTCGCCTTCT ACACCGACCT GATCGCGCTG 1500
CGGCGCGAGC TGCCGGTCGA TGCGCCGGCG CGCGAGGTGG ATGCCGACGA GGCGCGCGGC 1560
GTCTTCGCGT TCAGCCGCGG CCCGCTGCGG GTCACGGTCG CGCTGCGCCC CGGACCGGTC 1620
GGGGTGCCCG AGCACGGGGG CCTCGTGCTC GCCTACGGCG AGGTGCGCGC CGGCGCCGCC 1680
GGACTGCACC TCGACGGGCC GGGAGCCGCG ATCGTGCGCC TCGAG                  1725
```

(18)INFORMATION FOR SEQ ID NO:18:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:23
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:18:

GCSAACCGST GGTGGTGGGA CGT 23

(19)INFORMATION FOR SEQ ID NO:19:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:3252

(B)TYPE:nucleic acid

(C)strandedness:double

(D)TOPOLOGY:linear

(ii)MOLECULE TYPE:genomic DNA

(vi)ORIGINAL SOURCE:

(A)ORGANISM:Arthrobacter sp.

(B)INDIVIDUAL ISOLATE:S34 (FERM BP-6450)

(ix)FEATURE:

(A)NAME/KEY:5'UTR

(B)LOCATION:1..742

(C)IDENTIFICATION METHOD:E

(A)NAME/KEY:mat peptide

(B)LOCATION:743..3013

(C)IDENTIFICATION METHOD:E

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:19:

```
ATGCCGACGA CGAACTTGAG CGCGTTCTCG GGCACCCGCG AGAGCGGTCC GCGCACGGCG   60
GCGCCCAGTG CCACGACGAG CACGATCGCG GCGAGCGCCG CGACGACGGC GACCGGCAGG  120
CGCCCCTGAT TGCTGGCGAA GGTGAGCACG ATGAAGACCA CCTCGAGGCC CTCGAGCAAC  180
ACACCTTTGA ACGACACGGT GAACGCGTAC CAATCGGAGA CCCCGAACCG GCTCTCGCGC  240
CGGGCGCTCT CGGCCGCCTC GACCTGACGC CGGAAGGCAG CCTCCTCGTC ACGGAGAGCC  300
CTGCGCCCTG CCGCGCGCAG CACCGCCTTG CGCAGCCAGC CGAGCCCGAA GACGAGCAGC  360
AACCCGCCGA CGACGAGGCG CAGCACGGCC AGCGGCAGCA GCAGGATCGC GGGACCGACG  420
AGCGCGACGG CCGCGGCCAG CACCACCACG GCGACGGCGG CACCTGTCAG CGCCGACCGC  480
CAGCTGCGGG TGGCGCCGAC CGCGACGACG ATCGTGGTCG CCTCCACCGC CTCGACCACG  540
CAGGCGAGGA ACACGGCGGC GAACAGGGCG ACGGCGGTCA TCGGCCCAGC AGACGGTTGA  600
CCATCACGGC ACGCTAGCGC CATTGCTCAC AGGAAGGGCC AAGACGCCCG CAACGCGGCA  660
CCCGTGGACG GCGCGTACCG GCGTGTGACC GATCGTGTCA ACCGGTGGCG CCCGCCCCGA  720
GCACCTGCGT AGATTCGGCC TC GTG CCC GCC AGT ACC TAC CGC CTT CAG ATC    772
                          Met Pro Ala Ser Thr Tyr Arg Leu Gln Ile
                            1               5                  10

TCG GCG GAG TTC ACC CTC TTC GAC GCG GCG CGC ATC GTG CCC TAC CTG    820
Ser Ala Glu Phe Thr Leu Phe Asp Ala Ala Arg Ile Val Pro Tyr Leu
                15                  20                  25

CAC CGC CTC GGC GCC GAC TGG CTG TAC CTC TCG CCG CTG CTC GAG TCC    868
His Arg Leu Gly Ala Asp Trp Leu Tyr Leu Ser Pro Leu Leu Glu Ser
            30                  35                  40

GAG TCG GGC TCC TCG CAC GGC TAC GAC GTG GTC GAC CAC TCC CGC GTC    916
Glu Ser Gly Ser Ser His Gly Tyr Asp Val Val Asp His Ser Arg Val
        45                  50                  55

GAC GCC GCC CGC GGC GGG CCG GAG GGG CTC GCC GAG CTC TCC CGT GCG    964
Asp Ala Ala Arg Gly Gly Pro Glu Gly Leu Ala Glu Leu Ser Arg Ala
    60                  65                  70

GCG CAC GAG CGC GGC ATG GGC GTC GTC GTC GAC ATC GTG CCC AAC CAC   1012
Ala His Glu Arg Gly Met Gly Val Val Val Asp Ile Val Pro Asn His
75                  80                  85                  90

GTC GGC GTC GCG ACG CCG AAG GCG AAC CGC TGG TGG TGG GAC GTT CTG   1060
Val Gly Val Ala Thr Pro Lys Ala Asn Arg Trp Trp Trp Asp Val Leu
                95                 100                 105

GCC CGT GGA CAG CGG TCG GAG TAC GCC GAC TAC TTC GAC ATC GAC TGG   1108
Ala Arg Gly Gln Arg Ser Glu Tyr Ala Asp Tyr Phe Asp Ile Asp Trp
            110                 115                 120
```

```
GAG TTC GGC GGC GGC AGG CTG CGC CTG CCC GTG CTC GGC GAC GGC CCC   1156
Glu Phe Gly Gly Gly Arg Leu Arg Leu Pro Val Leu Gly Asp Gly Pro
        125                 130                 135
GAC GAG CTC GAC GCG CTG AGA GTG GAT GGC GAC GAG CTC GTC TAC TAC   1204
Asp Glu Leu Asp Ala Leu Arg Val Asp Gly Asp Glu Leu Val Tyr Tyr
    140                 145                 150
GAG CAC CGC TTC CCG ATC GCC GAG GGC ACC GGC GGC GGC ACC CCG CGC   1252
Glu His Arg Phe Pro Ile Ala Glu Gly Thr Gly Gly Gly Thr Pro Arg
155                 160                 165                 170
GAG GTG CAC GAC CGG CAG CAC TAC GAG CTG ATG TCG TGG CGG CGG GCC   1300
Glu Val His Asp Arg Gln His Tyr Glu Leu Met Ser Trp Arg Arg Ala
                175                 180                 185
GAC CAC GAC CTC AAC TAC CGC CGC TTC TTC GCC GTG AAC ACG CTC GCC   1348
Asp His Asp Leu Asn Tyr Arg Arg Phe Phe Ala Val Asn Thr Leu Ala
            190                 195                 200
GCC GTA CGC GTC GAA GAC CCG CGC GTG TTC GAC GAC ACC CAC CGC GAG   1396
Ala Val Arg Val Glu Asp Pro Arg Val Phe Asp Asp Thr His Arg Glu
        205                 210                 215
ATC GGC CGC TGG ATC GCC GAG GGC CTC GTC GAC GGC CTG CGC GTC GAC   1444
Ile Gly Arg Trp Ile Ala Glu Gly Leu Val Asp Gly Leu Arg Val Asp
    220                 225                 230
CAC CCC GAC GGG CTG CGC GCC CCC GGC GAC TAC CTG CGC CGT CTC GCC   1492
His Pro Asp Gly Leu Arg Ala Pro Gly Asp Tyr Leu Arg Arg Leu Ala
235                 240                 245                 250
GAG CTC GCC CAA GGC AGG CCG ATC TGG GTC GAG AAG ATC ATC GAG GGC   1540
Glu Leu Ala Gln Gly Arg Pro Ile Trp Val Glu Lys Ile Ile Glu Gly
                255                 260                 265
GAC GAG CGG ATG CCC CCG CAG TGG CCC ATC GCC GGC ACC ACC GGC TAC   1588
Asp Glu Arg Met Pro Pro Gln Trp Pro Ile Ala Gly Thr Thr Gly Tyr
            270                 275                 280
GAC GCG CTG GCC GGG ATC GAC CGG GTG CTC GTC GAC CCC GCG GGC GAG   1636
Asp Ala Leu Ala Gly Ile Asp Arg Val Leu Val Asp Pro Ala Gly Glu
        285                 290                 295
CAT CCG CTC ACC CAG ATC GTC GAC GAG GCG GCA GGC AGC CCC CGG CGC   1684
His Pro Leu Thr Gln Ile Val Asp Glu Ala Ala Gly Ser Pro Arg Arg
    300                 305                 310
TGG GCC GAG CTG GTT CCC GAG CGC AAG CGG GCC GTC GCC CGC GGC ATC   1732
Trp Ala Glu Leu Val Pro Glu Arg Lys Arg Ala Val Ala Arg Gly Ile
315                 320                 325                 330
CTG AAC TCC GAG ATC CGC CGC GTC GCC CGC GAA CTC GGA GAG GTC GCC   1780
Leu Asn Ser Glu Ile Arg Arg Val Ala Arg Glu Leu Gly Glu Val Ala
                335                 340                 345
GGC GAC GTC GAA GAC GCG CTC GTC GAG ATC GCC GCC GCC CTG TCC GTC   1828
Gly Asp Val Glu Asp Ala Leu Val Glu Ile Ala Ala Ala Leu Ser Val
            350                 355                 360
TAC CGC AGC TAC CTG CCG TTC GGG CGC GAG CAC CTC GAC GAA GCC GTG   1876
Tyr Arg Ser Tyr Leu Pro Phe Gly Arg Glu His Leu Asp Glu Ala Val
        365                 370                 375
GCC GCC GCG CAG GCC GCA GCC CCC CAG CTC GAG GCC GAC CTC GCC GCC   1924
Ala Ala Ala Gln Ala Ala Ala Pro Gln Leu Glu Ala Asp Leu Ala Ala
    380                 385                 390
GTC GGC GCA GCG CTC GCC GAC CCG GGC AAC CCC GCC GCG CTC CGC TTC   1972
Val Gly Ala Ala Leu Ala Asp Pro Gly Asn Pro Ala Ala Leu Arg Phe
395                 400                 405                 410
CAG CAG ACC AGC GGC ATG ATC ATG GCC AAG GGC GTC GAG GAC AAC GCG   2020
Gln Gln Thr Ser Gly Met Ile Met Ala Lys Gly Val Glu Asp Asn Ala
```

```
                    415                 420                 425
TTC TAC CGC TAC CCC CGG CTC ACC TCG CTG ACC GAG GTC GGG GGA GAC   2068
Phe Tyr Arg Tyr Pro Arg Leu Thr Ser Leu Thr Glu Val Gly Gly Asp
                430                 435                 440
CCG AGC CTG TTC GCG ATC GAC GCG GCC GCC TTC CAC GCG GCG CAG CGC   2116
Pro Ser Leu Phe Ala Ile Asp Ala Ala Ala Phe His Ala Ala Gln Arg
            445                 450                 455
GAC CGC GCC GCC CGG CTG CCC GAG TCG ATG ACG ACG CTG ACC ACC CAC   2164
Asp Arg Ala Ala Arg Leu Pro Glu Ser Met Thr Thr Leu Thr Thr His
        460                 465                 470
GAC ACC AAG CGC AGC GAA GAC ACC CGG GCG CGG ATC ACC GCG CTC GCC   2212
Asp Thr Lys Arg Ser Glu Asp Thr Arg Ala Arg Ile Thr Ala Leu Ala
475                 480                 485                 490
GAG GCC CCC GAA CGC TGG CGG CGC TTC CTG ACC GAG GTC GGC GGG CTC   2260
Glu Ala Pro Glu Arg Trp Arg Arg Phe Leu Thr Glu Val Gly Gly Leu
                495                 500                 505
ATC GGA ACG GGC GAC CGG GTG CTG GAG AAC CTG ATC TGG CAG GCG ATC   2308
Ile Gly Thr Gly Asp Arg Val Leu Glu Asn Leu Ile Trp Gln Ala Ile
            510                 515                 520
GTC GGC GCG TGG CCG GCG AGC CGG GAG CGG CTC GAG GCC TAC GCG CTG   2356
Val Gly Ala Trp Pro Ala Ser Arg Glu Arg Leu Glu Ala Tyr Ala Leu
        525                 530                 535
AAG GCC GCG CGC GAA GCC GGC GAG TCG ACC GAC TGG ATC GAC GGC GAC   2404
Lys Ala Ala Arg Glu Ala Gly Glu Ser Thr Asp Trp Ile Asp Gly Asp
    540                 545                 550
CCC GCG TTC GAA GAG CGG CTG ACC CGC CTG GTC ACG GTC GCC GTC GAG   2452
Pro Ala Phe Glu Glu Arg Leu Thr Arg Leu Val Thr Val Ala Val Glu
555                 560                 565                 570
GAG CCG CTC GTG CAC GAG CTG CTC GAG CGG CTC GTC GAC GAG CTG ACG   2500
Glu Pro Leu Val His Glu Leu Leu Glu Arg Leu Val Asp Glu Leu Thr
                575                 580                 585
GCG GCC GGG TAC TCC AAC GGC CTC GCG GCG AAG CTG CTG CAG CTG CTC   2548
Ala Ala Gly Tyr Ser Asn Gly Leu Ala Ala Lys Leu Leu Gln Leu Leu
            590                 595                 600
GCC CCC GGA ACC CCC GAC GTG TAC CAG GGC ACG GAA CGC TGG GAC CGG   2596
Ala Pro Gly Thr Pro Asp Val Tyr Gln Gly Thr Glu Arg Trp Asp Arg
        605                 610                 615
TCG CTG GTG GAC CCG GAC AAC CGT CGC CCC GTG GAT TTC GCC GCG GCA   2644
Ser Leu Val Asp Pro Asp Asn Arg Arg Pro Val Asp Phe Ala Ala Ala
    620                 625                 630
TCC GAG CTG CTC GAC CGC CTC GAC GGC GGC TGG CGG CCG CCC GTC GAC   2692
Ser Glu Leu Leu Asp Arg Leu Asp Gly Gly Trp Arg Pro Pro Val Asp
635                 640                 645                 650
GAG ACC GGC GCG GTC AAG ACG CTC GTC GTC TCC CGC GCG CTG CGG CTG   2740
Glu Thr Gly Ala Val Lys Thr Leu Val Val Ser Arg Ala Leu Arg Leu
                655                 660                 665
CGC CGC GAC CGG CCC GAG CTG TTC ACC GCG TAC CAC CCG GTC ACG GCG   2788
Arg Arg Asp Arg Pro Glu Leu Phe Thr Ala Tyr His Pro Val Thr Ala
            670                 675                 680
CGC GGC GCG CAG GCC GAG CAC CTG ATC GGC TTC GAC CGC GGC GGC GCG   2836
Arg Gly Ala Gln Ala Glu His Leu Ile Gly Phe Asp Arg Gly Gly Ala
        685                 690                 695
ATC GCC CTG GCC ACC CGC CTG CCG CTC GGC CTC GCC GCC GCA GGC GGC   2884
Ile Ala Leu Ala Thr Arg Leu Pro Leu Gly Leu Ala Ala Ala Gly Gly
    700                 705                 710
TGG GGC GAC ACG GTC GTC GAC GTC GGC GAG CGG AGC CTG CGC GAC GAG   2932
```

```
Trp Gly Asp Thr Val Val Asp Val Gly Glu Arg Ser Leu Arg Asp Glu
715                 720                 725                 730
CTG ACC GGC CGC GAG GCC CGC GGA GCG GCG CGC GTG GCC GAG TTG TTC   2980
Leu Thr Gly Arg Glu Ala Arg Gly Ala Ala Arg Val Ala Glu Leu Phe
                735                 740                 745
GCC GAC TAC CCC GTC GCC CTG CTG GTG GAG ACA TGAACCGACG ATTCCCGGTC 3033
Ala Asp Tyr Pro Val Ala Leu Leu Val Glu Thr
            750                 755
TGGGCGCCCC AGGCCGCGCA GGTGACGCTC GTCGTGGGCC AAGGCCGCGC CGAACTCCCG 3093
CTGACCCGCG ACGAGAACGG ATGGTGGGCT CTTCAGCAGC CGTGGGACGG CGGCCCCGAC 3153
CTCGTCGACT ACGGCTACCT CGTCGACGGC AAGGGCCCCT TCGCCGACCC GCGGTCGCTG 3213
CGGCAGCCGC GCGGCGTGCA CGAGCTCGGC CGCGAATTC                        3252
```

(20)INFORMATION FOR SEQ ID NO:20:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:26

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:20:

ATGCCCGCCA GTACCTACCG CCTTCA

(21)INFORMATION FOR SEQ ID NO:21:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:25

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:21:

TCATGTCTCC ACCAGCAGGG CGACG

(22)INFORMATION FOR SEQ ID NO:22:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:50

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:22:

AATTCTTTTT TAATAAAATC AGGAGGAATC TAGATGTTTA CTAGTCTGCA

(23)INFORMATION FOR SEQ ID NO:23:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:42

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:23:

GACTAGTAAA CATCTAGATT CCTCCTGATT TTATTAAAAA AG

(24)INFORMATION FOR SEQ ID NO:24:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:33
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:24:

AAATCTAGAT GCCCGCCAGT ACCTACCGCC TTC 33

(25)INFORMATION FOR SEQ ID NO:25:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:33
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:25:

AAAACTAGTT TATCATGTCT CCACCAGCAG GGC 33

(26)INFORMATION FOR SEQ ID NO:26:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:22
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:26:

ATCGGTGATG TCGGCGATAT AG 22

(27)INFORMATION FOR SEQ ID NO:27:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:29
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:27:

GTACTGGCGG GCATATTTTT TCCTCCTGA 29

(28)INFORMATION FOR SEQ ID NO:28:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:31
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:28:

AATCAGGAGG AAAAAATATG CCCGCCAGTA C 31

(29)INFORMATION FOR SEQ ID NO:29:

(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:22
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:29:

TCGACGATCT GGGTGAGCGG AT 22

(30)INFORMATION FOR SEQ ID NO:30:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:22

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:30:

TCGACGAGCA CCCGGTCGAT CC 22

(31)INFORMATION FOR SEQ ID NO:31:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:26

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:31:

CARTGGGAYG AYGAYGTNCA YCAYGC 26

B1

(32)INFORMATION FOR SEQ ID NO:32:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:2218

(B)TYPE:nucleic acid

(C)strandedness:double

(D)TOPOLOGY:linear

(ii)MOLECULE TYPE:genomic DNA

(vi)ORIGINAL SOURCE:

(A)ORGANISM:Arthrobacter sp.

(B)INDIVIDUAL ISOLATE:S34 (FERM BP-6450)

(ix)FEATURE:

(A)NAME/KEY:mat peptide

(B)LOCATION:477..2201

(C)IDENTIFICATION METHOD:E

(A)NAME/KEY:3S'UTR

(B)LOCATION:2202..2218

(C)IDENTIFICATION METHOD:E

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:32:

```
CTGCAGCTGC TCGCCCCCGG AACCCCCGAC GTGTACCAGG GCACGGAACG CTGGGACCGG    60
TCGCTGGTGG ACCCGGACAA CCGTCGCCCC GTGGATTTCG CCGCGGCATC CGAGCTGCTC   120
GACCGCCTCG ACGGCGGCTG GCGGCCGCCC GTCGACGAGA CCGGCGCGGT CAAGACGCTC   180
GTCGTCTCCC GCGCGCTGCG GCTGCGCCGC GACCGGCCCG AGCTGTTCAC CGCGTACCAC   240
CCGGTCACGG CGCGCGGCGC GCAGGCCGAG CACCTGATCG GCTTCGACCG CGGCGGCGCG   300
ATCGCCCTGG CCACCCGCCT GCCGCTCGGC CTCGCCGCCG CAGGCGGCTG GGGCGACACG   360
GTCGTCGACG TCGGCGAGCG GAGCCTGCGC GACGAGCTGA CCGGCCGCGA GGCCCGCGGA   420
GCGGCGCGCG TGGCCGAGTT GTTCGCCGAC TACCCCGTCG CCCTGCTGGT GGAGAC ATG   479
                                                                Met
                                                                 1

AAC CGA CGA TTC CCG GTC TGG GCG CCC CAG GCC GCG CAG GTG ACG CTC     527
Asn Arg Arg Phe Pro Val Trp Ala Pro Gln Ala Ala Gln Val Thr Leu
          5                  10                  15

GTC GTG GGC CAA GGC CGC GCC GAA CTC CCG CTG ACC CGC GAC GAG AAC     575
Val Val Gly Gln Gly Arg Ala Glu Leu Pro Leu Thr Arg Asp Glu Asn
     20                  25                  30
```

```
GGA TGG TGG GCT CTT CAG CAG CCG TGG GAC GGC GGC CCC GAC CTC GTC     623
Gly Trp Trp Ala Leu Gln Gln Pro Trp Asp Gly Gly Pro Asp Leu Val
     35                  40                  45
GAC TAC GGC TAC CTC GTC GAC GGC AAG GGC CCC TTC GCC GAC CCG CGG     671
Asp Tyr Gly Tyr Leu Val Asp Gly Lys Gly Pro Phe Ala Asp Pro Arg
 50                  55                  60                  65
TCG CTG CGG CAG CCG CGC GGC GTG CAC GAG CTC GGC CGC GAA TTC GAC     719
Ser Leu Arg Gln Pro Arg Gly Val His Glu Leu Gly Arg Glu Phe Asp
                 70                  75                  80
CCC GCC CGC TAC GCG TGG GGC GAC GAC GGA TGG CGC GGC CGA GAC CTC     767
Pro Ala Arg Tyr Ala Trp Gly Asp Asp Gly Trp Arg Gly Arg Asp Leu
             85                  90                  95
ACC GGA GCC GTG ATC TAC GAA CTG CAC GTC GGC ACC TTC ACC CCT GAG     815
Thr Gly Ala Val Ile Tyr Glu Leu His Val Gly Thr Phe Thr Pro Glu
        100                 105                 110
GGA ACG CTG GAC AGC GCC ATC CGT CGC CTC GAC CAC CTG GTG CGC CTC     863
Gly Thr Leu Asp Ser Ala Ile Arg Arg Leu Asp His Leu Val Arg Leu
    115                 120                 125
GGC GTC GAC GCG GTC GAG CTG CTG CCC GTC AAC GCG TTC AAC GGC ACC     911
Gly Val Asp Ala Val Glu Leu Leu Pro Val Asn Ala Phe Asn Gly Thr
130                 135                 140                 145
CAC GGC TGG GGC TAC GAC GGG GTG CTC TGG TAC GCG GTG CAC GAG CCC     959
His Gly Trp Gly Tyr Asp Gly Val Leu Trp Tyr Ala Val His Glu Pro
                150                 155                 160
TAC GGC GGC CCG GAG GCG TAC CAG CGC TTC GTC GAC GCC TGC CAC GCC    1007
Tyr Gly Gly Pro Glu Ala Tyr Gln Arg Phe Val Asp Ala Cys His Ala
            165                 170                 175
CGC GGC CTC GCC GTC GTG CAG GAC GTC GTC TAC AAC CAC CTG GGC CCG    1055
Arg Gly Leu Ala Val Val Gln Asp Val Val Tyr Asn His Leu Gly Pro
        180                 185                 190
AGC GGC AAC CAC CTG CCC GAC TTC GGC CCC TAC CTC GGG TCG GGC GCC    1103
Ser Gly Asn His Leu Pro Asp Phe Gly Pro Tyr Leu Gly Ser Gly Ala
    195                 200                 205
GCC AAC ACC TGG GGC GAC GCG CTG AAC CTC GAC GGG CCG CTC TCC GAC    1151
Ala Asn Thr Trp Gly Asp Ala Leu Asn Leu Asp Gly Pro Leu Ser Asp
210                 215                 220                 225
GAG GTG CGG CGG TAC ATC ATC GAC AAC GCG GTG TAC TGG CTG CGC GAC    1199
Glu Val Arg Arg Tyr Ile Ile Asp Asn Ala Val Tyr Trp Leu Arg Asp
                230                 235                 240
ATG CAC GCC GAC GGG CTG CGG CTC GAC GCC GTG CAC GCG CTG CGC GAC    1247
Met His Ala Asp Gly Leu Arg Leu Asp Ala Val His Ala Leu Arg Asp
            245                 250                 255
GCC CGC GCG CTG CAC CTG CTC GAA GAG CTC GCC GCC CGC GTC GAC GAG    1295
Ala Arg Ala Leu His Leu Leu Glu Glu Leu Ala Ala Arg Val Asp Glu
        260                 265                 270
CTG GCG GGC GAG CTC GGC CGG CCG CTG ACG CTC ATC GCC GAG AGC GAC    1343
Leu Ala Gly Glu Leu Gly Arg Pro Leu Thr Leu Ile Ala Glu Ser Asp
    275                 280                 285
CTG AAC GAC CCG AAG CTG ATC CGC TCC CGC GCG GCG CAC GGC TAC GGC    1391
Leu Asn Asp Pro Lys Leu Ile Arg Ser Arg Ala Ala His Gly Tyr Gly
290                 295                 300                 305
CTC GAC GCC CAG TGG GAC GAC GAC GTG CAC CAC GCG GTG CAC GCC AAC    1439
Leu Asp Ala Gln Trp Asp Asp Asp Val His His Ala Val His Ala Asn
                310                 315                 320
GTG ACC GGC GAG ACC GTC GGC TAC TAC GCC GAC TTC GGC GGG CTC GGC    1487
Val Thr Gly Glu Thr Val Gly Tyr Tyr Ala Asp Phe Gly Gly Leu Gly
```

```
                325                 330                 335
GCC CTC GTC AAG GTG TTC CAG CGC GGC TGG TTC CAC GAC GGC ACC TGG   1535
Ala Leu Val Lys Val Phe Gln Arg Gly Trp Phe His Asp Gly Thr Trp
        340                 345                 350
TCG AGC TTC CGC GAG CGG CAC CAC GGC CGG CCG CTC GAC CCC GAC ATC   1583
Ser Ser Phe Arg Glu Arg His His Gly Arg Pro Leu Asp Pro Asp Ile
    355                 360                 365
CCG TTC CGC CGG CTC GTC GCC TTC GCG CAG GAT CAC GAC CAG GTC GGC   1631
Pro Phe Arg Arg Leu Val Ala Phe Ala Gln Asp His Asp Gln Val Gly
370                 375                 380                 385
AAC CGA GCG GTC GGC GAC CGC ATG TCG GCG CAG GTC GGC GAG GGT TCG   1679
Asn Arg Ala Val Gly Asp Arg Met Ser Ala Gln Val Gly Glu Gly Ser
                390                 395                 400
CTC GCC GCC GCG GCG GCG CTC GTG CTG CTC GGC CCG TTC ACC CCG ATG   1727
Leu Ala Ala Ala Ala Ala Leu Val Leu Leu Gly Pro Phe Thr Pro Met
            405                 410                 415
CTG TTC ATG GGC GAG GAG TGG GGC GCG CGC ACC CCG TGG CAG TTC TTC   1775
Leu Phe Met Gly Glu Glu Trp Gly Ala Arg Thr Pro Trp Gln Phe Phe
        420                 425                 430
ACC TCC CAC CCC GAG CCC GAG CTG GGG GAG GCG ACG GCG CGC GGG CGC   1823
Thr Ser His Pro Glu Pro Glu Leu Gly Glu Ala Thr Ala Arg Gly Arg
    435                 440                 445
ATC GCC GAG TTC GCC CGC ATG GGC TGG GAC CCG GCA GTC GTG CCC GAC   1871
Ile Ala Glu Phe Ala Arg Met Gly Trp Asp Pro Ala Val Val Pro Asp
450                 455                 460                 465
CCG CAG GAC CCG GCC ACC TTC GCC CGC TCG CAC CTG GAC TGG TCC GAG   1919
Pro Asp Asp Pro Ala Thr Phe Ala Arg Ser His Leu Asp Trp Ser Glu
                470                 475                 480
CCC GAG CGG GAA CCG CAC GCG GGC CTG CTC GCC TTC TAC ACC GAC CTG   1967
Pro Glu Arg Glu Pro His Ala Gly Leu Leu Ala Phe Tyr Thr Asp Leu
            485                 490                 495
ATC GCG CTG CGG CGC GAG CTG CCG GTC GAT GCG CCG GCG CGC GAG GTG   2015
Ile Ala Leu Arg Arg Glu Leu Pro Val Asp Ala Pro Ala Arg Glu Val
        500                 505                 510
GAT GCC GAC GAG GCG CGC GGC GTC TTC GCG TTC AGC CGC GGC CCG CTG   2063
Asp Ala Asp Glu Ala Arg Gly Val Phe Ala Phe Ser Arg Gly Pro Leu
    515                 520                 525
CGG GTC ACG GTC GCG CTG CGC CCC GGA CCG GTC GGG GTG CCC GAG CAC   2111
Arg Val Thr Val Ala Leu Arg Pro Gly Pro Val Gly Val Pro Glu His
530                 535                 540                 545
GGG GGC CTC GTG CTC GCC TAC GGC GAG GTG CGC GCC GGC GCC GCC GGA   2159
Gly Gly Leu Val Leu Ala Tyr Gly Glu Val Arg Ala Gly Ala Ala Gly
                550                 555                 560
CTG CAC CTC GAC GGG CCG GGA GCC GCG ATC GTG CGC CTC GAG           2201
Leu His Leu Asp Gly Pro Gly Ala Ala Ile Val Arg Leu Glu
            565                 570                 575
TGACGCGGCT GGGTACC                                               2218
```

(33)INFORMATION FOR SEQ ID NO:33:

(i)SEQUENCE CHARACTERISTICS:

(A)LENGTH:25

(B)TYPE:nucleic acid

(C)strandedness:single

(D)TOPOLOGY:linear

(xi)SEQUENCE DESCRIPTION:SEQ ID NO:33:

ATGAACCGAC GATTCCCGGT CTGGG 25

(34)INFORMATION FOR SEQ ID NO:34:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:25
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:34:

TCACTCGAGG CGCACGATCG CGGCT 25

(35)INFORMATION FOR SEQ ID NO:35:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:36
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:35:

AAATCTAGAT GAACCGACGA TTCCCGGTCT GGGCGC 36

(36)INFORMATION FOR SEQ ID NO:36:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:36
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:36:

AAAACTAGTT TATCACTCGA GGCGCACGAT CGCGGC 36

(37)INFORMATION FOR SEQ ID NO:37:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:28
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:37:

ATCGTCGGTT CATATTTTTT CCTCCTGA 28

(38)INFORMATION FOR SEQ ID NO:38:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:28
(B)TYPE:nucleic acid
(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:38:

AATCAGGAGG AAAAAATATG AACCGACG 28

(39)INFORMATION FOR SEQ ID NO:39:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH:22
(B)TYPE:nucleic acid

(C)strandedness:single
(D)TOPOLOGY:linear
(xi)SEQUENCE DESCRIPTION:SEQ ID NO:39:

AGGTGGTTGT AGACGACGTC CT

22